

Published in final edited form as:

Birth Defects Res C Embryo Today. 2009 June ; 87(2): 123–130. doi:10.1002/bdrc.20152.

Building Developmental Gene Regulatory Networks

Enhu Li¹ and Eric H. Davidson^{1,*}

¹Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

Animal development is an elaborate process programmed by genomic regulatory instructions. Regulatory genes encode transcription factors and signal molecules, and their expression is under the control of cis-regulatory modules that define the logic of transcriptional responses to the inputs of other regulatory genes. The functional linkages amongst regulatory genes constitute the gene regulatory networks (GRNs) that govern cell specification and patterning in development. Constructing such networks requires identification of the regulatory genes involved and characterization of their temporal and spatial expression patterns. Interactions (activation/repression) among transcription factors or signals can be investigated by large-scale perturbation analysis, in which the function of each gene is specifically blocked. Resultant expression changes are then integrated to identify direct linkages, and to reveal the structure of the GRN. Predicted GRN linkages can be tested and verified by cis-regulatory analysis. The explanatory power of the GRN was shown in the lineage specification of sea urchin endomesoderm. Acquiring such networks is essential for a systematic and mechanistic understanding of the developmental process.

Keywords

gene regulatory network; development; cis-regulatory analysis; sea urchin

Introduction

Gene regulatory networks (GRNs) causally relate the genomic regulatory sequence to the dynamic developmental process (Davidson, 2006). A GRN includes mainly regulatory genes encoding transcription factors and signaling molecules, and most importantly it makes explicit the instructions for spatial and temporal expression of regulatory genes embedded in their cis-regulatory modules. These are patches of DNA sequence that include the various elements recognized and bound by transcription factors, including those that respond to signal transduction pathways. Cis-regulatory modules determine the outputs of the genes they service, i.e., the time, place, and rate of expression, according to their inputs, which are provided by the transcription factors presented in the current cellular regulatory state (Istrail et al., 2007). In developmental GRNs, the outputs of regulatory genes mainly feed into the cis-regulatory modules controlling other regulatory genes, except at the periphery of the GRN where the linkages lead to differentiation genes and structural morphogenesis genes. Typically GRNs include complicated feedback, feedforward, and cross-regulatory loops.

GRNs offer a mechanistic view of how the embryonic body plan is formed through lineage specification, differentiation, and morphogenesis. They directly relate the processes of development to the genomic regulatory code, and thus provide a central avenue of

*Correspondence to: Eric H. Davidson, Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. Davidson@caltech.edu.

understanding why development occurs as it does. In this article we briefly review the means used to construct GRNs, taking the sea urchin embryo GRNs as a model, and discuss briefly some of their characteristics.

Building GRNs based on functional experimental data

Solving a GRN related to any development process requires knowledge of what transcription factors and signal molecules are involved in the system, when and where the genes are expressed, and most importantly, how the regulatory genes interact with one another. None of these data is simply acquired, and extensive measurements are typically needed to characterize the temporal and spatial expression of all relevant regulatory genes, while a huge number of data points is required to identify the genuine linkages among interacting genes as the activities of individual genes are alternately removed from the system one by one.

The first step in building a GRN is to identify the regulatory genes involved. When the complete genome sequence is available, the most comprehensive and arguably the best solution is a top-down approach, i.e., a genome-wide survey of all predicted regulatory genes followed by characterization of their spatial and temporal expression patterns. The gene repertoire used for developmental control is relatively conserved, and it is amazing that despite evolutionary divergence and the distinct organizations of their body plans, all Bilateria retain a relatively similar tool kit of regulatory genes (Erwin and Davidson, 2002). Thus, transcription factors and signal molecules encoded by the regulatory genome can be identified by sequence-based homology searching, an approach that was exhaustively applied to the genome of the sea urchin (*Strongylocentrotus purpuratus*) (Cameron et al., 2004; Howard-Ashby et al., 2006a, Howard-Ashby et al., 2006b; Materna et al., 2006; Rizzo et al., 2006; Tu et al., 2006). Sequences of known regulatory gene families can be used as queries for BLAST or profile searches; or alternatively, DNA-binding motifs can be used as search features to identify potential transcription factors.

The putative regulatory genes then need to be characterized with respect to their expression, to determine when and where in the developmental process they are utilized. The initial problem is to categorize regulatory genes by the phase of development in which they are expressed. Microarray analysis (high-density tiling array, or cDNA array) is suited for genome-wide transcription profiling, and is in common use as a high-throughput method (e.g., for the sea urchin mRNA transcriptome, see Samanta et al., 2006; Wei et al., 2006). However, for measurements of the quantities of transcripts of up to several hundred interesting regulatory genes at different times or in different regulatory conditions, the requirement we are here concerned with, we prefer quantitative PCR (QPCR). This method offers greater sensitivity, far more accuracy, and larger dynamic range (Howard-Ashby et al., 2006a, Howard-Ashby et al., 2006b; Materna et al., 2006; Tu et al., 2006). Recently we have found a new technology, the NanoString “nCounter Analysis System”, extremely useful for quantitative expression studies. NanoString allows direct simultaneous measurement of mRNA sequence levels for 50–500 genes, by counting the number of “barcoded” fluorescent antisense RNA probes hybridized to their respective targets. NanoString measurements are directly comparable to QPCR measurements in terms of sensitivity, but have a much higher throughput (Geiss et al., 2008; Su et al., 2009).

Once genes pertinent to a certain developmental stage have been identified through temporal expression profiling, the next question is where they are expressed. The most straightforward way to determine this in sea urchin embryos is by whole mount *in situ* hybridization, using chromogenic probes specific to the gene of interest (Ransick et al., 1993; Arenas-Mena et al., 2000; Ransick, 2004). It is of course necessary that the

developmental scheme, the cell lineage or embryonic anatomy, and the specification map of the organism be known *a priori*, so that patterns revealed by *in situ* hybridization of the analyzed gene can be matched by inspection to a particular functional developmental territory. Here the extremely well known and relatively simply constructed sea urchin embryo provides a very considerable advantage. Each territory expresses a unique set of regulatory genes, i.e., at each time a unique regulatory state. All territory specific gene expression, and thus all developmental process, depends causally on the regulatory state (Davidson, 2006). Ubiquitously expressed genes contribute less to the specification of a particular territory; they may provide at times essential cis-regulatory boosters to given genes, but cannot in themselves determine the qualitative diversification of spatial domains. Therefore they are generally considered low in priority when constructing a GRN. Mapping regulatory gene expression to various territorial domains is essential for subsequent analysis of the GRN *per se*, but this is a time-consuming and challenging task compared to temporal analysis, as practical and reliable *in situ* analysis methods available for spatial determination are still limited to relatively low-throughput studies of a few genes at a time.

Difficult though it may be, a close to complete description of expression of the regulatory gene set, at high temporal and spatial resolution, is an underlying requirement of a relevant GRN model. Though a primitive GRN built upon a limited number of (key) nodes is sometimes illuminating, the GRN becomes far more accurate and potent as it approaches completeness. "Completeness" for developmental GRNs is defined as incorporation of all specifically expressed regulatory genes that contribute to the territorial regulatory state (Levine and Davidson, 2005; Davidson, 2006). Incomplete GRNs lack the power to predict genuine network connections, and whenever networks are missing key nodes, they are inevitably rich in false positives. The ultimate goal of constructing the GRN is comprehensiveness, incorporation of all related components, capture of all genuine interactions, and inclusion of all stages of the developmental process.

Perturbation methods and analysis tools

The key information on which the GRN model is formulated is identification of epistatic relations among the regulatory genes, including both activation and repression. The method is to perturb the system and examine the responses, and thereby infer the interactions among network components.

The most precise and by far most extensive type of perturbation used in construction of the sea urchin GRNs has been introduction of morpholino-substituted antisense oligonucleotides (MASOs), although in particular circumstances gene expression patterns are altered in other ways as well. From the beginning of the effort to construct the sea urchin embryo GRN, MASO blockade of gene expression has been the major tool (Davidson et al., 2002a, Davidson et al., 2002b). MASOs base pair stably and specifically to target mRNAs, thus blocking their translation or splicing. Their embryonic toxicity is low, and they can be delivered by microinjection into the egg. Disrupting expression of a regulatory gene has various effects, depending on the importance of the network linkage that is interrupted. Development can be severely impaired if operation of a key node is blocked, but in many cases the changes to phenotype are marginal. More generally, multiple regulatory genes always contribute to any given morphological feature, and interference with expression of any one of them is likely to produce the same phenotype (e.g., see Oliveri et al., 2008). Phenotype is thus a very poor index of the actual network linkages, and the effects of the perturbation on regulatory gene expression can only be determined by gene specific assays. Spatial expansion or reduction in the territorial expression of target genes can be investigated by *in situ* hybridization, while change in expression levels can be measured with various quantification tools (e.g. QPCR, or the NanoString nCounter). Quantitative

assays are at present far more suited for large-scale analysis. In sea urchin developmental studies, because of the natural variability of different batches of eggs, as well as experimental variation, if an upstream perturbation causes a reproducible > three-fold change of expression as measured by QPCR, it is considered significant. This is a conservative threshold, and recently we have been able to reduce the threshold of significance to two-fold or less by use of the NanoString nCounter.

Establishing network linkages from perturbation data

Perturbation assays followed by expression profiling reveals the significant expression changes caused directly as well as indirectly by disruption of network linkages. As a general point, this information, together with constraints available from data on time and place of gene expression, and from knowledge of the biology of the process, suffices for the construction of a GRN model. The GRN model should be explanatory of, and compatible with, the observed data, as well as with any previously identified interactions. Some level of parsimony should be achieved so that the large amount of observed interactions can be interpreted as a minimal number of linkages. That is, parallel and redundant inputs are not assumed unless specifically indicated, so that if an input could be indirect it is assumed that it is. A brief summary of the logic rules we use for interpretation of perturbation data and convolution of expression with perturbation data is listed in Table 1.

Linkages proposed to be direct must be consistent with the constraints arising from comparison of the temporal and spatial expression of the putative upstream and downstream genes. Since the driver gene product must reach a functional concentration, it must be expressed earlier and not later than its putative target genes. However, because of the relatively slow default turnover rates of macromolecules in sea urchin embryos, transcriptional events occurring hours earlier may still display effects on downstream genes hours later (Bolouri and Davidson, 2003). At some point, however, the driver and its target genes must meet the logic restriction that both be expressed in the same spatial domain, or in overlapping domains. In some cases, a significant decrease in the level of a transcription factor results in expressional changes of other genes transcribed in adjacent but non-overlapping domains. This kind of result directly implies the existence of signal(s) that communicate between the regulatory states of the two domains: the driver gene is upstream of the signaling ligand, and the responding genes are the (direct or indirect) targets of the signal transduction pathway activated in the recipient cells. Inductive signaling is not uncommon, but rather an essential and almost universal aspect of development; signaling is the major mechanism by which spatial regulatory state is progressively changed during embryogenesis (Davidson, 2006). Signaling is not only used to induce new regulatory states in the cells of an adjacent territory, but also within territories (the “community effect”; Gurdon, 1987; Davidson 2006; Nam et al., 2007). Here it plays the role of ensuring that the individual cells of a territory express the same regulatory state at the same level. This again has to be incorporated in the architecture of a prospective GRN model. The interplay between spatial and temporal expression and perturbation results extends to repression as well. Target genes repressed by another transcription factor should have an exclusive territorial pattern with respect to the upstream gene, or they can both be expressed in the same domain, but at successive developmental stages.

The putative linkages surviving these logic gates should be further examined to trim out possible indirect interactions. Experience shows that direct targets are surprisingly easy to identify *a priori*. In many cases, the response to a MASO perturbation of specific gene expression is broad, and many targets show expression changes at various levels due to direct plus indirect linkages. The magnitude of expression changes turns out to provide a useful clue, particularly for genes with multiple inputs or outputs. Higher-fold changes of

expression are indicative of direct activation or repression, while weaker effects in the same cells are usually indirect. Thus if an indirect path can be found in a GRN connecting the perturbed gene and a putative indirect target, the effect is indeed probably indirect. In accordance with this guideline, network building should start with those linkage pairs with highest expression changes.

Though a primary tool in identifying linkages among regulatory genes, quantitative assessment of the consequences of perturbation treatments is not the only useful kind of measurement, and there are often situations where additional methods/strategies are absolutely necessary for solving network structure. For instance, if the expression domain of a driver gene covers only a fraction of that of its target gene, the resultant quantitative change of expression of the downstream gene upon knockdown of this driver input is expected to be marginal and the linkage between might easily be missed. In this kind of situation, further evidence such as *in situ* hybridization of the target gene in the MASO-treated embryos is needed. Similarly the indirect but real relationships of feedforward subcircuits are better solved by employing rescue experiments, and/or double MASO assays.

Authentication of the GRN by cis-regulatory measurements

The predicted direct relationships, or linkages, between driver and target genes in a GRN model are subject to immediate and specific validation by cis-regulatory analysis. In biochemical terms, driver gene products (transcription factors) should be able to bind to the regulatory region of the target gene, promoting or inhibiting its transcription. Therefore, analysis of the relevant cis-regulatory modules of the target genes should reveal binding site(s) for the protein products of the driver genes, and if the interaction is direct, mutagenesis of the binding sites should abolish the regulatory effects exerted by the upstream genes. If not supported by the cis-regulatory analysis, the predicted linkages have to be revamped. Thus building GRNs is a reiterative process of prediction and authentication. The outcome is a GRN which is, as we say, cast in cis-regulatory concrete.

Network construction by perturbation, and authentication by cis-regulatory analysis are interdependent routes toward a complete GRN model. Because it is time consuming, cis-regulatory analysis has so far been valuable for confirmation of predicted interactions, rather than as a primary tool for GRN construction. However, this is going to change soon as new, very fast, high throughput methods for cis-regulatory analysis now being developed in our lab come online (unpublished data). It is important to note that cis-regulatory analysis may contribute more than mere confirmation or refutation of predicted GRN linkages: additional inputs are thus frequently identified, and in extensively studied examples, cis-regulatory analysis can yield otherwise inaccessible insights into the fundamental logic of gene regulation, as seen in the studies of the *endo16* and *Otx* genes of the sea urchin embryo (Yuh et al., 2001; Yuh et al., 2002).

In experimental systems, such as the sea urchin embryo, where the cis-regulatory control systems are immediately accessible to direct, functional experimental characterization in gene transfer experiments, the logic of validation is as follows: (i) the network model established in perturbation assays provides precise predictions of cis-regulatory input identities and functions at each GRN node; (ii) cis-regulatory modules which recreate the relevant phase of expression of any gene(s) in the network are tested to see if they respond as predicted to the MASO's which affect the endogenous gene; (iii) the predicted target sites in these cis-regulatory modules are identified and mutated to determine if this has the same effect as MASO treatment, assessed quantitatively and/or spatially using transgenic expression constructs. Thus direct GRN linkages can be unequivocally validated. In the current sea urchin embryo endomesoderm GRN shown in Figure 1, every thick line

represents a linkage that has been experimentally validated. Most of the important nodes of the network are thus authenticated, and soon all will be (or corrected; for always current versions of the GRN, <http://sugp.caltech.edu/endomes/#BioTapestryViewer>). The literature contains many attempts to build GRNs from chromatin immunoprecipitation and/or largely computational predictions of target sites, in experimental systems where cis-regulatory analysis is (or is believed to be) difficult. These models are in a different class as they are not generated from functional data, in contrast both to the initial perturbation-based GRN models discussed above and the validated cis-regulatory versions of such models. Or, said another way, if a complete list of regulatory players is known, and perturbation and cis-regulatory analysis are experimentally feasible, these together provide the necessary and sufficient bases for GRN construction and validation.

Predictive quality of network circuitry

In the process of completing a GRN, it sometimes happens that the structure of a subnetwork logically implies an additional missing node which mediates regulation between two nodes whose interaction is indirect. A good example of this was encountered in the course of solving the portion of the GRN which pertains to the skeletogenic lineages of the sea urchin embryo (the lavender domain on the left of Fig. 1). Initial studies of the micromere GRN showed that the transcription factor encoded by the gene *pmar1* is an early determinant, which is both necessary and sufficient for skeletogenic specification; indeed ectopic expression of *pmar1* suffices to cause the expressing cells to embark on skeletogenic specification (Oliveri et al., 2002; Oliveri et al., 2003). However, *pmar1* encodes a transcriptional repressor, not an activator, and yet its expression was seen clearly to be required for activation of a set of immediately downstream genes, including *delta*, *tbr*, *ets1*, and *alx*. The logical implication of this was that there must be a missing node in the GRN where there resides a gene encoding a second repressor subject to *pmar1* repression, so that the mechanism of activation of the downstream genes is actually a double negative gate (see Fig. 2). We provisionally named the predicted gene “Repressor of Micromeres”. Its predicted properties were sufficiently precise that by using them it was soon uncovered by a systematic search. First, early expressed regulatory genes were screened for those down regulated by the global expression of *pmar1* mRNA, as the GRN predicted that Repressor of Micromeres should be turned off when *pmar1* is over-expressed. The most promising candidate was the *hesC* gene, which encodes a canonical repressor. It had the predicted unusual pattern of expression, which was zygotic expression everywhere in the embryo except in the skeletogenic micromere lineage where *pmar1* is expressed. Further work proved that *hesC* is indeed the predicted Repressor of Micromeres (Revilla-i-Domingo et al, 2007; Oliveri et al, 2008). For example, treatment with *hesC* MASO resulted in the same phenotype embryo as that caused by global expression of *pmar1*, in which the whole embryo turns into mesenchymal cells expressing skeletogenic genes. Thus the target genes of the double negative gate, for instance *delta* and *alx1*, are expressed globally, just the same, either when *hesC* translation is blocked, or when *pmar1* is expressed all over the egg. Absent the initial predictive network analysis, the existence of the double negative gate would never have been revealed, least of all by traditional one-gene-at-a-time biology.

BioTapestry and its usage in constructing and illustrating the GRN

Considering the inherent complexity of GRNs, it is a challenging task to present GRN models to users; additionally, as more data are accumulated in high throughput analyses, computational tools are needed for processing these data, and incorporating them into GRN models. To this end, Biotapestry was developed as a free, open-source, and platform-independent software suite (<http://www.biotapestry.org/>). BioTapestry offers a toolset for

creating, editing, and illustrating GRNs (Longabaugh et al., 2005). The image of the sea urchin GRN in Figure 1 was built in BioTapestry.

The BioTapestry network is organized such that subnetworks from various domains of the embryo are represented horizontally in the diverse color coded blocks, while vertically are various developmental stages, time proceeding in the downward direction. The GRN shown in Figure 1 is a comprehensive view: it includes all components, and all linkages from all types of cells. It is termed the View from All Nuclei (VfA). This is the genomicist's view, as its purpose is to explicitly indicate all the inputs into each relevant cis-regulatory sequence. The VfA directly relates the A's, C's, G's, and T's of the regulatory genome to the architecture or topology of the network (Longabaugh et al, 2005;Davidson, 2006). The VfA also serves to visualize the modular subcircuits of which the GRN is composed (Levine and Davidson, 2005;Davidson, 2006). This is a bird's eye view, global but static. The developmental biologist's view, the View from the Nucleus (VfN), is more helpful in tracing the dynamics of cell specification. VfNs display the network linkages active in each cell type at each time. The succession of VfNs as development proceeds show the control functions causal to the initiation and progression of regulatory states.

A simplified example of a VfN is shown in Figure 3. This particular BioTapestry diagram shows how the neighboring skeletogenic (SM) and non-skeletogenic (NSM) lineages adopt different fates through the expression of distinct sets of regulatory genes. The SM lineage soon turns on *delta* downstream of the double negative gate as above. Delta is a signaling ligand which activates the Notch receptor of the adjacent cells. These cells adopt an NSM fate, an initial trigger for which is expression of genes responsive to the Delta-Notch pathway. In addition to the signaling mechanism, the distinct fates of SM and NSM are secured through a mutually exclusive repression system which we call the "exclusion function" (Oliveri et al, 2008; for simplicity only one component of this system is here illustrated). In the NSM cells, *gcm*, a direct target gene of the Delta-Notch pathway, represses the SM specific regulator *alx*; while in SM cells, *gcm* is similarly repressed (Fig. 3).

The kinetics with which GRNs operate lie outside the confines of this brief discussion. However, it is essential to note that the inter-connective topology of a BioTapestry GRN offers the in-depth understanding of the pathways required to understand the dynamic processes of development. For treatment of network kinetics with special relevance to the well developed sea urchin GRNs, see Bolouri and Davidson (2003); Longabaugh and Bolouri, (2006); Davidson and Levine, (2008a); Ben-Tabou de-Leon and Davidson, (2009). The kinetics can be solved knowing the network topology (and some necessary parameters), but the reverse is not true: the network topology can never be derived from the kinetics except for topologies that are atypically simple, unlike any of those in the GRN in Figure 1.

Concluding Comment

We have attempted in this article to summarize the kinds of experimental information thought processes required for synthetic construction of developmental GRNs from functional laboratory data. Unlike traditional bioscience focused on one or several genes at a time, GRN models rest upon the premise of systems biology that only at the level of complexity required to incorporate all or most of the moving parts of a process, can that process actually be causally explained. This shift in explanatory focus represents a fundamental reorientation in developmental biology (Davidson, 2008). Although the evidence and examples we discuss are strongly concentrated on the sea urchin GRN, similar systematic approaches are being applied to many other developing systems, both embryonic and postembryonic. Within the last few months, for example, two prominent collections of

papers have appeared devoted entirely to developmental GRNs in diverse animal systems (Cameron and Oliveri (eds), 2008; Davidson and Levine (eds), 2008b). This is a new and rapidly expanding field. The basic reason for the intense interest in GRNs is that for the first time, GRNs enable us to explicitly see how the genome encodes developmental process, why development occurs as it does, and what is the logic of the developmental program. Identification of more GRNs from various genetic systems is also of great importance because of the evolutionary insights that will accrue (Erwin and Davidson, 2009; Hinman and Davidson, 2007; Woodland and Zorn, 2008). The diverse body plans of different animals are formed by diverse GRNs, and the evolution of body plans is in mechanistic terms the evolution of GRNs.

Acknowledgments

Research was supported by NIH grants HD-37105 and GM-61005. Most of the funding for the purchase of the NanoString nCounter was a gift from the Ahmanson Research & Equipment Fund.

REFERENCES

- Arenas-Mena C, Cameron AR, Davidson EH. Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. *Development*. 2000; 127:4631–4643. [PubMed: 11023866]
- Ben-Tabou de-Leon S, Davidson EH. Gene regulation: gene control network in development. *Annu Rev Biophys Biomol Struct*. 2007; 36:191. [PubMed: 17291181]
- Ben-Tabou de-Leon S, Davidson EH. Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Dev Biol*. 2009; 325:317–328. [PubMed: 19028486]
- Bolouri H, Davidson EH. Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proc Natl Acad Sci USA*. 2003; 100:9371–9376. [PubMed: 12883007]
- Cameron RA, Rast JP, Brown CT. Genomic resources for the study of sea urchin development. *Methods Cell Biol*. 2004; 74:733–757. [PubMed: 15575629]
- Cameron RA, Oliveri P. Gene regulatory mechanisms. *Biochim Biophys Acta*. 2009; 1789:247–374.
- Davidson EH, Rast JP, Oliveri P, et al. *Science*. 2002a; 295:1669–1678. [PubMed: 11872831]
- Davidson EH, Rast JP, Oliveri P, et al. *Dev Biol*. 2002b; 246:162–190. [PubMed: 12027441]
- Davidson, EH. *The regulatory genome: gene regulatory networks in development and evolution*. 1st edition. Academic Press/Elsevier; San Diego: 2006.
- Davidson EH. *Developmental biology at the systems level*. *Biochim Biophys Acta*. 2008
- Davidson EH, Levine MS. Properties of developmental gene regulatory networks. *Proc Natl Acad Sci USA*. 2008a; 105:20063–20066. [PubMed: 19104053]
- Davidson EH, Levine MS. Gene networks in development and evolution. *Proc Natl Acad Sci USA*. 2008b; 105:20063–20117. [PubMed: 19104053]
- Erwin DH, Davidson EH. The last common bilaterian ancestor. *Development*. 2002; 129:3021–3032. [PubMed: 12070079]
- Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet*. 2009; 10:141–148. [PubMed: 19139764]
- Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008; 26:317–325. [PubMed: 18278033]
- Gurdon JB. Embryonic induction--molecular prospects. *Development*. 1987; 99:285–306. [PubMed: 3308408]
- Hinman VF, Davidson EH. Evolutionary plasticity of developmental gene regulatory network architecture. *Proc Natl Acad Sci USA*. 2007; 104:19404–19409. [PubMed: 18042699]
- Howard-Ashby M, Materna SC, Brown CT, et al. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. *Dev Biol*. 2006a; 300:90–107. [PubMed: 17054934]

- Howard-Ashby M, Materna SC, Brown CT, et al. Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development. *Dev Biol.* 2006b; 300:74–89. [PubMed: 17055477]
- Istrail S, De-Leon SB, Davidson EH. The regulatory genome and the computer. *Dev Biol.* 2007; 310:187–195. [PubMed: 17822690]
- Levine M, Davidson EH. Gene regulatory networks for development. *Proc Natl Acad Sci USA.* 2005; 102:4936–4942. [PubMed: 15788537]
- Longabaugh WJ, Davidson EH, Bolouri H. Computational representation of developmental genetic regulatory networks. *Dev Biol.* 2005; 283:1–16. [PubMed: 15907831]
- Longabaugh W, Bolouri H. Understanding the Dynamic Behavior of Genetic Regulatory Networks by Functional Decomposition. *Curr Genomics.* 2006; 7:333–341. [PubMed: 18079985]
- Materna SC, Howard-Ashby M, Gray RF, Davidson EH. The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. *Dev Biol.* 2006; 300:108–120. [PubMed: 16997293]
- Nam J, Su YH, Lee PY, et al. Cis-regulatory control of the nodal gene, initiator of the sea urchin oral ectoderm gene network. *Dev Biol.* 2007; 306:860–869. [PubMed: 17451671]
- Oliveri P, Carrick DM, Davidson EH. A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev Biol.* 2002; 246:209–228. [PubMed: 12027443]
- Oliveri P, Davidson EH, McClay DR. Activation of *pmar1* controls specification of micromeres in the sea urchin embryo. *Dev Biol.* 2003; 258:32–43. [PubMed: 12781680]
- Oliveri P, Tu Q, Davidson EH. Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci USA.* 2008; 105:5955–5962. [PubMed: 18413610]
- Ransick A, Ernst S, Britten RJ, Davidson EH. Whole mount in situ hybridization shows *Endo 16* to be a marker for the vegetal plate territory in sea urchin embryos. *Mech Dev.* 1993; 42:117–124. [PubMed: 8217840]
- Ransick A. Detection of mRNA by in situ hybridization and RT-PCR. *Methods Cell Biol.* 2004; 74:601–620. [PubMed: 15575623]
- Revilla-i-Domingo R, Oliveri P, Davidson EH. A missing link in the sea urchin embryo gene regulatory network: *hesC* and the double-negative specification of micromeres. *Proc Natl Acad Sci USA.* 2007; 104:12383–12388. [PubMed: 17636127]
- Rizzo F, Fernandez-Serra M, Squarzone P, et al. Identification and developmental expression of the *ets* gene family in the sea urchin (*Strongylocentrotus purpuratus*). *Dev Biol.* 2006; 300:35–48. [PubMed: 16997294]
- Samanta MP, Tongprasit W, Istrail S, et al. The transcriptome of the sea urchin embryo. *Science.* 2006; 314:960–962. [PubMed: 17095694]
- Smith J, Davidson EH. Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo. *Proc Natl Acad Sci USA.* 2008; 105:20089–20094. [PubMed: 19104065]
- Su YH, Li E, Geiss GK, et al. A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo. *Dev Biol.* 2009; 329:410–421. [PubMed: 19268450]
- Tu Q, Brown CT, Davidson EH, Oliveri P. Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev Biol.* 2006; 300:49–62. [PubMed: 17081512]
- Wei Z, Angerer RC, Angerer LM. A database of mRNA expression patterns for the sea urchin embryo. *Dev Biol.* 2006; 300:476–484. [PubMed: 17007833]
- Woodland HR, Zorn AM. The core endodermal gene network of vertebrates: combining developmental precision with evolutionary flexibility. *Bioessays.* 2008; 30:757–765. [PubMed: 18623060]
- Yuh CH, Bolouri H, Davidson EH. Cis-regulatory logic in the *endo16* gene: switching from a specification to a differentiation mode of control. *Development.* 2001; 128:617–629. [PubMed: 11171388]
- Yuh CH, Brown CT, Livi CB, et al. Patchy interspecific sequence similarities efficiently identify positive cis-regulatory elements in the sea urchin. *Dev Biol.* 2002; 246:148–161. [PubMed: 12027440]

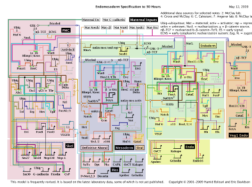


Figure 1.

Gene regulatory network (GRN) of sea urchin endomesoderm specification up to 30hrs. Sea urchin development proceeds from top to bottom, and various domains of embryos are represented with color-coded blocks, including skeletogenic mesoderm (SM) cells (lavendar), non-skeletogenic mesoderm (NSM) (purple), and endoderm [veg1 endoderm (orange) and endoderm derived from veg2 endomesoderm precursors (yellow)]. Maternal factors are boxed and reside in the top gray block. The current GRN sea urchin endomesoderm is available online and constantly updated (<http://sugp.caltech.edu/endomes/#EndomesNetwork>). The model is built on procedures described in this article, including expression profiling and spatial characterization of regulatory genes, and large scale perturbation analysis. A large portion of network linkages has been confirmed by cis-regulatory analysis (thick lines). Detailed descriptions and reviews of the GRN can be found elsewhere (Davidson, 2006; Ben-Tabou de-Leon and Davidson, 2007; Oliveri et al., 2008; Smith and Davidson, 2008).

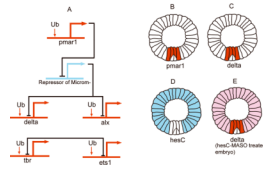


Figure 2.

Identification of predicted “Repressor of Micromeres” as the *hesC* gene, and the double negative gate by which specification of the sea urchin skeletogenic micromere lineage occurs. (A) GRN subcircuit that includes the *pmar1* double negative gate. This model was built as described in Oliveri et al. (2002). A putative repressor gene (Repressor of Micromeres) was predicted as the target of *pmar1*, the role of which is to repress micromere lineage skeletogenic genes such as *alx1*, *delta*, *ets1*, and *tbr* globally, except where *pmar1* is active. (B) Diagram of sea urchin embryo with *pmar1* expressed in micromere (vegetal-most cells in red). (C) Expression domain of *delta* gene. Its spatial expression is identical to that of *pmar1* and expected to be under the control of *pmar1* through the double negative gate. (D) Expression territory of *hesC*. *hesC* was identified as the Repressor of Micromeres, and expressed in regions exclusive to *pmar1* or *delta*. (E) Expression of *delta* in embryos treated with *hesC* MASO. Knock-down of *hesC* causes ectopic *delta* gene expression in regions (pink) extending outside of the micromeres (red). On, off: gene is actively transcribed, or silenced. Ub, ubiquitous factors.

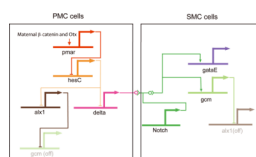


Figure 3.

Simplified version of endomesoderm subnetworks of SM and NSM, as views from the nucleus (VfN's). (A) SM subnetwork; (B) NSM subnetwork. Network components are extracted from Fig. 1, and only some of the linkages are shown. Symbols of protein-DNA or protein-protein interactions: \longrightarrow , transcriptional activation; $\longrightarrow|$ transcriptional repression; $\longrightarrow o \longrightarrow$, ligand-receptor binding; $\longrightarrow o$ —, signal transduction process.

TABLE 1
Rules for Building Gene Regulatory Networks Based on Perturbation Data

Temporal restriction
a) Direct transcriptional activation requires coincident expression of driver and target or driver expression earlier than targets
b) Direct activation by ligand-binding in the recipient cells requires that the downstream genes not be expressed prior to the ligand
Spatial restriction
a) Direct transcriptional activation possible when driver and target are co-expressed in same domain
b) Indirect interaction through signal(s) is implied if effect is observed in the domain in which the perturbed gene is not expressed
c) Direct transcriptional repression requires that the repressor and its targets not overlap in their expression domain, or in the same domain at successive developmental stages
Parsimonious topology
a) Direct target of a transcription factor is more strongly affected than indirect targets
b) A given interaction is presumed indirect if identified direct linkages suffice to explain the data
Special linkages
Double negative control logic is required when interference with a canonical repressor causes decrease in expression of target genes